

WHAT IS CLAIMED IS:

1. A method of confirming that a candidate genomic region harbors a gene associated with a detectable trait comprising the steps of:

5       constructing a candidate region distribution of test values using a plurality of biallelic markers in a candidate genomic region suspected of harboring said gene associated with said detectable trait, said candidate region distribution of test values being indicative of the difference in the frequencies of said plurality of biallelic markers in said candidate region in individuals who possess said detectable trait and control individuals who do not possess said detectable trait;

10       constructing a random region distribution of test values using a plurality of biallelic markers in random genomic regions which are not suspected of harboring said gene associated with said detectable trait, said random region distribution of test values being indicative of the difference in the frequencies of said plurality of biallelic markers in said random genomic regions in individuals who possess said detectable trait and control individuals who do not possess said detectable trait; and

15       determining whether said candidate region distribution of test values and said random region distribution of test values are significantly different from one another.

20       2. The method of Claim 1, wherein said step of constructing a candidate region distribution of test values comprises performing a haplotype analysis on each possible combination of biallelic markers in each group in a series of groups of biallelic markers in said candidate region, calculating test values for each possible combination, and including the test value for the haplotype which has the greatest association with said trait in said candidate region distribution of test values for each group in said series of groups of biallelic markers in said candidate genomic region and wherein said step of constructing a random region distribution of test values comprises performing a haplotype analysis on each possible combination of biallelic markers in each group in a series of groups of biallelic markers in said random genomic regions, calculating test values for each possible combination, and including the test value for the haplotype which has the greatest association with said trait in said random region distribution of test values for each group in said series of groups of biallelic markers in said random genomic regions.

25       3. The method of Claim 2, wherein said steps of performing a haplotype analysis on each possible combination of biallelic markers in each group in said series of groups of biallelic markers in said candidate genomic region and calculating said test values for each combination comprises the steps of:

calculating the frequencies for each combination of biallelic markers in each group in said series of groups of biallelic markers in said candidate genomic region in individuals expressing said detectable trait;

calculating the frequencies for each combination of biallelic markers in each group in said series of groups of biallelic markers in said candidate genomic region in individuals who do not express said detectable trait; and

comparing the haplotype frequencies in individuals who express said trait and individuals who do not express said trait by performing a chi-squared analysis to yield said test values.

4. The method of Claim 3, wherein said steps of performing a haplotype analysis on each possible combination of biallelic markers in each group in said series of groups of biallelic markers in said random genomic regions and calculating said test values for each combination comprises the steps of:

calculating the frequencies for each combination of biallelic markers in each group in said series of groups of biallelic markers in said random genomic regions in individuals expressing said detectable trait;

calculating the frequencies for each combination of biallelic markers in each group in said series of groups of biallelic markers in said random genomic regions in individuals in individuals who do not express said detectable trait; and

comparing the haplotype frequencies in individuals who express said trait and individuals who do not express said trait by performing a chi-squared analysis to yield said test values.

5. The method of Claim 4, wherein said step of comparing said candidate region distribution of test values to said random region distribution of test values comprises performing a Wilcoxon rank test.

6. The method of Claim 4, wherein said step of comparing said candidate region distribution of test values to said random region distribution of test values comprises performing a Kolmogorov-Smirnov test.

7. The method of Claim 4, said step of comparing said candidate region distribution of test values to said random region distribution of test values comprises performing both a Wilcoxon rank test and a Kolmogorov-Smirnov test.

8. The method of Claim 4, wherein each of said groups of biallelic markers in said series of groups of biallelic markers in said candidate genomic region and each of said groups of biallelic markers in said series of groups of biallelic markers in said random genomic regions comprises 3 biallelic markers.

9. The method of Claim 4, wherein each of said groups of biallelic markers in said series of groups of biallelic markers in said candidate genomic region and each of said groups of biallelic markers in said series of groups of biallelic markers in said random genomic regions comprises at least 3 biallelic markers.

5 10. The method of Claim 4, wherein said biallelic markers in each of said groups in said series of groups of biallelic markers in said candidate genomic region have an average intermarker distance selected from the group consisting of one marker every 3kb, one marker every 5kb, one marker every 10kb, one marker every 20kb, and one marker every 30kb.

10 11. The method of Claim 10, wherein said biallelic markers in each of said groups in said series of groups of biallelic markers in said random genomic regions have an average intermarker distance selected from the group consisting of one marker every 3kb, one marker every 5kb, one marker every 10kb, one marker every 20kb, and one marker every 30kb.

12. The method of Claim 4 further comprising selecting random genomic regions for use in said haplotype analysis which have at least 3 biallelic markers therein.

15 13. The method of Claim 12, further comprising selecting random genomic regions for use in said haplotype analysis in which said biallelic markers have an average intermarker distance sufficient conducting a haplotype analysis.

20 14. The method of Claim 13 further comprising selecting random genomic regions for use in said haplotype analysis wherein said at least 3 biallelic markers are in Hardy-Weinberg equilibrium in individuals expressing said detectable trait and control individuals who do not express said detectable trait.

25 15. The method of Claim 14 further comprising selecting random genomic regions for use in said haplotype analysis in which said at least 3 biallelic markers are not in complete linkage disequilibrium to be useful in conducting a haplotype analysis.

30 16. The method of Claim 3 further comprising selecting biallelic markers in said candidate genomic region which are in Hardy-Weinberg equilibrium in individuals expressing said detectable trait and control individuals who do not express said detectable trait for use in said haplotype analysis.

17. The method of Claim 16 further comprising determining the total number of markers in said candidate genomic region.

18. The method of Claim 4 further comprising the step of verifying that the biallelic markers in said random genomic regions are appropriate for use in the haplotype analysis by:

randomly dividing said biallelic markers in said random genomic regions into a first verification group and a second verification group, wherein said first verification group and said second verification group contain a substantially identical number of biallelic markers;

constructing a first verification distribution of test values for the biallelic markers in said first verification group by performing a haplotype analysis on each possible combination of biallelic markers in each group in a series of groups of biallelic markers in said first verification group, calculating test values for each possible combination, and including the test value for the haplotype which has the greatest association with said trait in said first verification distribution of test values for each group in said series of groups of biallelic markers in said first verification group;

constructing a second verification distribution of test values for the biallelic markers in said second verification group by performing a haplotype analysis on each possible combination of biallelic markers in each group in a series of groups of biallelic markers in said second verification group, calculating test values for each possible combination, and including the test value for the haplotype which has the greatest association with said trait in said second verification distribution of test values for each group in said series of groups of biallelic markers in said second verification group;

determining whether said first verification distribution and said second verification distribution are significantly different from one another, wherein said biallelic markers in said random genomic regions are appropriate for use in the haplotype analysis if said first verification distribution and said second verification distribution are not significantly different from one another.

19. The method of Claim 18 wherein said steps of performing a haplotype analysis on each possible combination of biallelic markers in each group in said series of groups of biallelic markers in said first and second verification groups and calculating said test values for each combination comprises the steps of:

calculating the frequencies for each combination of biallelic markers in said first verification group in each group in said series of groups of biallelic markers in individuals expressing said detectable trait;

calculating the frequencies for each combination of biallelic markers in said first verification group in each group in said series of groups of biallelic markers in individuals who do not express said detectable trait;

comparing the haplotype frequencies of said biallelic markers in said first verification group in individuals who express said trait and individuals who do not express said trait by performing a chi-squared analysis to yield said test values;

calculating the frequencies for each combination of biallelic markers in said second verification group in each group in said series of groups of biallelic markers in individuals expressing said detectable trait;

calculating the frequencies for each combination of biallelic markers in said second verification group in each group in said series of groups of biallelic markers in individuals who do not express said detectable trait;

comparing the haplotype frequencies of said biallelic markers in said second verification group in individuals who express said trait and individuals who do not express said trait by performing a chi-squared analysis to yield said test values

20. The method of Claim 19, wherein said step of determining whether said first verification distribution and said second verification distribution are significantly different from one another comprises performing a Wilcoxon rank test on said first and second verification distributions.

21. The method of Claim 19, wherein said step of determining whether said first verification distribution and said second verification distribution are significantly different from one another comprises performing a Kolmogorov-Smirnov test on said first and second verification distributions.

22. The method of Claim 19, wherein said step of determining whether said first verification distribution and said second verification distribution are significantly different from one another comprises performing a both a Kolmogorov-Smirnov test and a Wilcoxon rank test on said first and second verification distributions.

23. The method of Claim 19, wherein each of said groups of biallelic markers in said series of groups of biallelic markers in said first verification group and each of said groups of biallelic markers in said series of groups of biallelic markers in said second verification group contains 3 biallelic markers.

24. The method of Claim 19, wherein each of said groups of biallelic markers in said series of groups of biallelic markers in said first verification group and each of said groups of biallelic markers in said series of groups of biallelic markers in said second verification group contains more than 3 biallelic markers.

25. The method of Claim 1, wherein said method is performed by a computer.

26. The method of Claim 25, wherein said computer provides an output indicative of whether said candidate region distribution of test values and said random region distribution of test values are significantly different.

27. The method of Claim 26 further comprising further evaluating said candidate genomic region to identify candidate genes which might be associated with said detectable trait

if said output indicates that said candidate region distribution of test values and said random region distribution of test values are significantly different.

28. The method of Claim 1 further comprising further evaluating said candidate genomic region to identify candidate genes which might be associated with said detectable trait if said candidate region distribution of test values and said random region distribution of test values are significantly different.

29. The method of Claim 4 wherein the frequencies for each combination of biallelic markers in each group in said series of groups of biallelic markers in said candidate genomic region and in said random genomic regions in individuals expressing said detectable trait are calculated using the Expectation Maximization algorithm; and

the frequencies for each combination of biallelic markers in each group in said series of groups of biallelic markers in said candidate genomic region and said random genomic regions in individuals who do not express said detectable trait are calculated using the Expectation Maximization algorithm.

30. A method of determining whether a candidate genomic region harbors a gene associated with a detectable trait comprising determining whether the association of a plurality of biallelic markers located in said candidate genomic region with said detectable trait is significantly different than the association of a plurality of biallelic markers located in a plurality of random genomic regions.

31. The method of Claim 30 wherein the determination of whether the association of said plurality of biallelic markers located in said candidate genomic region with said detectable trait is significantly different than the association of said plurality of biallelic markers located in a plurality of random genomic regions comprises:

constructing a candidate region distribution of test values using said biallelic markers in said candidate genomic region, said candidate region distribution of test values being indicative of the difference in the haplotype frequencies of said biallelic markers in said candidate region in individuals who possess said detectable trait and control individuals who do not possess said detectable trait;

constructing a random region distribution of test values using said biallelic markers in said genomic region said random region distribution of test values being indicative of the difference in the haplotype frequencies of said biallelic markers in said random genomic regions in individuals who possess said detectable trait and control individuals who do not possess said detectable trait; and

comparing said candidate region distribution of test values with said random region distribution of test values.

32. The method of Claim 31, wherein said step of constructing a candidate region distribution of test values comprises performing a haplotype analysis on each possible combination of biallelic markers in each group in a series of groups of biallelic markers in said candidate region, calculating test values for each possible combination, and including the test value for the haplotype which has the greatest association with said trait in said candidate region distribution of test values for each group in said series of groups of biallelic markers in said candidate genomic region and wherein said step of constructing a random region distribution of test values comprises performing a haplotype analysis on each possible combination of biallelic markers in each group in a series of groups of biallelic markers in said random genomic regions, calculating test values for each possible combination, and including the test value for the haplotype which has the greatest association with said trait in said random region distribution of test values for each group in said series of groups of biallelic markers in said random genomic regions.

33. A computer system for confirming that a candidate genomic region harbors a gene associated with a detectable trait, wherein the computer system comprises instructions that when executed perform the method of:

constructing a candidate region distribution of test values using a plurality of biallelic markers in a candidate genomic region suspected of harboring said gene associated with said detectable trait, said candidate region distribution of test values being indicative of the difference in the frequencies of said plurality of biallelic markers in said candidate region in individuals who possess said detectable trait and control individuals who do not possess said detectable trait;

constructing a random region distribution of test values using a plurality of biallelic markers in random genomic regions, said random region distribution of test values being indicative of the difference in the frequencies of said plurality of biallelic markers in said random genomic regions in individuals who possess said detectable trait and control individuals who do not possess said detectable trait; and

determining whether said candidate region distribution of test values and said random region distribution of test values are significantly different from one another.

34. The computer system of Claim 33, wherein said instructions for constructing a candidate region distribution of test values comprise instructions for performing a haplotype analysis on each possible combination of biallelic markers in each group in a series of groups of biallelic markers in said candidate region, calculating test values for each possible combination, and including the test value for the haplotype which has the greatest association with said trait in said candidate region distribution of test values for each group in said series of groups of

biallelic markers in said candidate genomic region and wherein said instructions for constructing a random region distribution of test values comprise instructions for performing a haplotype analysis on each possible combination of biallelic markers in each group in a series of groups of biallelic markers in said random genomic regions, calculating test values for each possible combination, and including the test value for the haplotype which has the greatest association with said trait in said random region distribution of test values for each group in said series of groups of biallelic markers in said random genomic regions.

35. The computer system of Claim 34, wherein said instructions for performing a haplotype analysis on each possible combination of biallelic markers in each group in said series of groups of biallelic markers in said candidate genomic region and calculating said test values for each combination comprise instructions for:

calculating the frequencies for each combination of biallelic markers in each group in said series of groups of biallelic markers in said candidate genomic region in individuals expressing said detectable trait;

calculating the frequencies for each combination of biallelic markers in each group in said series of groups of biallelic markers in said candidate genomic region in individuals who do not express said detectable trait; and

comparing the haplotype frequencies in individuals who express said trait and individuals who do not express said trait by performing a chi-squared analysis to yield said test values.

36. The computer system of Claim 35, wherein said instructions for performing a haplotype analysis on each possible combination of biallelic markers in each group in said series of groups of biallelic markers in said random genomic regions and calculating said test values for each combination comprise instructions for:

calculating the frequencies for each combination of biallelic markers in each group in said series of groups of biallelic markers in said random genomic regions in individuals expressing said detectable trait;

calculating the frequencies for each combination of biallelic markers in each group in said series of groups of biallelic markers in said random genomic regions in individuals in individuals who do not express said detectable trait; and

comparing the haplotype frequencies in individuals who express said trait and individuals who do not express said trait by performing a chi-squared analysis to yield said test values.





possible combination, and including the test value for the haplotype which has the greatest association with said trait in said random region distribution of test values for each group in said series of groups of biallelic markers in said random genomic regions.

42. The programmed storage device of Claim 41, wherein said instructions for performing a haplotype analysis on each possible combination of biallelic markers in each group in said series of groups of biallelic markers in said candidate genomic region and calculating said test values for each combination comprise instructions for:

calculating the frequencies for each combination of biallelic markers in each group in said series of groups of biallelic markers in said candidate genomic region in individuals expressing said detectable trait;

calculating the frequencies for each combination of biallelic markers in each group in said series of groups of biallelic markers in said candidate genomic region in individuals who do not express said detectable trait; and

comparing the haplotype frequencies in individuals who express said trait and individuals who do not express said trait by performing a chi-squared analysis to yield said test values.

43. The programmed storage device of Claim 42, wherein said instructions for performing a haplotype analysis on each possible combination of biallelic markers in each group in said series of groups of biallelic markers in said random genomic regions and calculating said test values for each combination comprise instructions for:

calculating the frequencies for each combination of biallelic markers in each group in said series of groups of biallelic markers in said random genomic regions in individuals expressing said detectable trait;

calculating the frequencies for each combination of biallelic markers in each group in said series of groups of biallelic markers in said random genomic regions in individuals in individuals who do not express said detectable trait; and

comparing the haplotype frequencies in individuals who express said trait and individuals who do not express said trait by performing a chi-squared analysis to yield said test values.

44. The programmed storage device of Claim 43, wherein said instructions for comparing said candidate region distribution of test values to said random region distribution of test values comprise instructions for performing a Wilcoxon rank test.

45. The programmed storage device of Claim 43, wherein said instructions for comparing said candidate region distribution of test values to said random region distribution of test values comprise instructions for performing a Kolmogorov-Smirnov test.

46. The programmed storage device of Claim 43, wherein said instructions for comparing said candidate region distribution of test values to said random region distribution of test values comprise instructions for performing both a Wilcoxon rank test and a Kolmogorov-Smirnov test.

5 47. The programmed storage device of Claim 40, wherein said programmed storage device is selected from the group consisting of a hard disk, a floppy disk, Random Access Memory, Read Only Memory and Electrically Erasable Programable Read Only Memory.

0988289-061001